A COMPARISON OF SOME TEST STATISTICS OF THE KOLMOGOROV TYPE

by

Jasper Paul Hendren

United States
Naval Postgraduate School

THESIS

A COMPARISON OF SOME TEST STATISTICS
OF THE KOLMOGOROV TYPE

by

Jasper Paul Hendren

October 1969

A Comparison of Some Test Statistics

of the Kolmogorov Type

by

Jasper Paul Hendren
Lieutenant, United States Navy
B.A., Hanover College, 1964

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
October 1969

ABSTRACT

It is natural to associate optimal estimation and optimal statistical testing. In this paper continuous function estimation of the cumulative distribution is used to define two test statistics that compete with the Kolmogorov D_N statistic. The first statistic, C_N, is attributed to Pyke and the second, R_N, is obtained by polygonalizing the sample distribution function. It is known that both are asymptotically equivalent to the Kolmogorov statistic. Using the methods of J. Durbin, the small sample distributions are tabled as well as the critical points for significance levels of .20, .10, .05, .025, and .01. It is shown that R_N is stochastically smaller than D_N, and it appears that C_N is also smaller than D_N.

TABLE OF CONTENTS

LIST OF DRAWINGS

ACKNOWLEDGEMENT

The author wishes to express his thanks to Professor Robert R. Read for the guidance he gave during the course of this investigation. In addition, his thoughtful critique of the work accounted for many improvements in the final written form and organization of this thesis.

I. INTRODUCTION

The popular method for testing the simple hypothesis H_o: $F = F_o$, where F is the unknown cumulative distribution of the population sampled and F_o is the proposed continuous distribution of the sample, is the Kolmogorov D_N statistic [Ref. 4]. The asymptotic distribution of D_N was developed by Kolmogorov [Ref. 2] and tabled by Smirnov [Ref. 6]; the small sample distribution was tabled by Massey [Refs. 3 and 4] and has been treated recently (among others) by Durbin [Ref. 1]. A modification of the Kolmogorov test has been proposed by Pyke [Ref. 5] and the distribution of this test, called C_N, can be obtained from Durbin [Ref. 1].

The Kolmogorov D_N statistic is associated with a step function estimation of F in the computation of the test statistic. This function, however, is from a vacuous class under the assumptions of the method, since a priorily F is assumed to be continuous. Of course the estimator converges to the continuous function F as the sample size becomes large without limit.

The question of estimating the distribution F is separate from testing H_o. However, it is natural to associate the two and to expect optimal estimation to yield optimal testing methods. Two of the statistics considered in this paper may be viewed as being based upon continuous estimators of F. It is shown that each is asymptotically equivalent to D_N.

The statistics investigated are presented in section two. The asymptotic nature of each and the relation between each is shown in section three. Section four presents the computational method used in tabling the distribution of the statistics introduced in section two. Final comparisons are covered in section four. The tables are in the appendices.

II. BASIC DEFINITIONS, KOLMOGOROV STATISTIC, PYKE STATISTIC, AND A PROPOSED STATISTIC

Let $Y_1, Y_2, \ldots, Y_N$ be a random sample from a continuous population with cumulative distribution function $F(y)$. The methods presented below are for testing the simple hypothesis

$$H_o : F = F_o \quad .$$

The procedures discussed are based on the order statistics of the sample where

$$-\infty < X_1 \leq X_2 \leq \ldots \leq X_N < \infty$$

are the ordered sample points, and the sample cumulative distribution is

$$F_N(x) = \begin{cases} 0 & X < X_1 \\ j/N & X_j \leq X < X_{j+1} \\ 1 & X > X_N \end{cases} \quad .$$

The Kolmogorov D_N statistic for testing H_o is

$$D_N = \sup_x |F_N(x) - F_o(x)|$$

and has the property of being distribution free under H_o [Ref. 4]. This property allows for computation of the cumulative distribution of D_N to be done with F_o assumed to be uniform (0,1) without loss of generality.

Lemma 1 : $D_N = \max_{1 \le j \le N} \{|X_j - (j/N)|, |X_j - (j-1)/N|\}$.

Proof: After transformation under H_o the statistic becomes

$$D_N = \sup_{0 \le x \le 1} |X - F_N(x)| \quad .$$

The assertion is obvious from Figure one.

A modification of the Kolmogorov statistic is attributed to Pyke [Ref. 5] where the test statistic is

$$C_N = \max_{1 \le j \le N} |X_j - j/(n+1)| \quad .$$

Implicit in this method is the idea that the j^{th} order statistic estimates the ((j+1)/N)(100) percentile of F. Such estimators are unbiased when F_o is uniform (0,1).

The statistic C_N may be viewed as being based on a continuous function estimator of F constructed by connecting the points (0,0); $(X_1, 1/(N+1))$;...; $(X_N, N/(N+1))$; (1,1) with straight line segments. This is apparent since any point on the line segment L from $(X_j, j/(N+1))$ to $(X_{j+1}, (j+1)/(n+1))$ is a convex combination of the endpoints and hence the maximum separation between points on the line Y = X and L must occur at one of the endpoints of L.

It is of interest to consider a sample cumulative distribution formed by connecting the points (0,0); $(X_1, 1/N)$;...; $(X_N, 1)$ with straight line segments and the resultant statistic

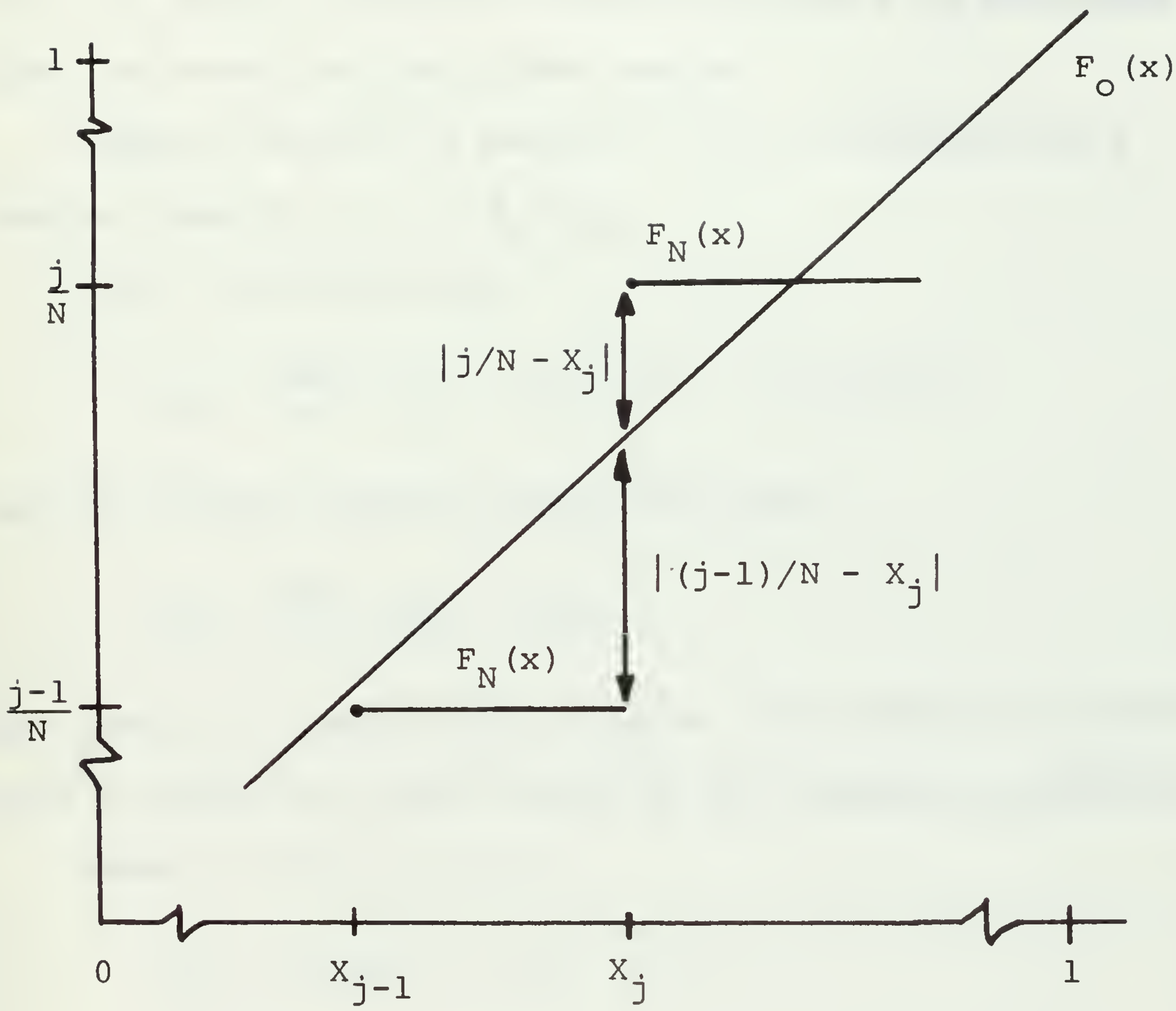

Figure 1. Lemma 1

$$R_N = \sup_x |S_N(x) - F_o(x)|$$

where $S_N(x)$ is the sample cumulative distribution modified as described above. Properties of R_N are presented in section three and its distribution for finite N is tabled in Appendix B.

III. ASYMPTOTIC EQUIVALENCE OF THE TEST STATISTICS

It is well known that

$$\lim_{N\to\infty} P[\sqrt{N}\, D_N > Z] = 2 \sum_{r=1}^{\infty} (-1)^{r-1} \exp\{-2r^2 Z^2\}$$

[Ref. 2] and it is shown below that both $\sqrt{N}\, R_N$ and $\sqrt{N}\, C_N$ have the same limiting distribution.

Lemma 2: For all N and all F R_N is stochastically smaller than D_N, i.e., $R_N \leq D_N$.

Proof: By lemma one

$$D_N = \max_j \{|X_j - j/N|, |X_j - (j-1)/N|\}$$

and it follows from the definition that

$$R_N = \max_j \{|X_j - j/N|\} \quad .$$

The result is immediately obvious. The latter is maximized over a subset of those for which the former is maximized.

Lemma 3: For all $\varepsilon > 0$

$$\lim_{N\to\infty} P[\sqrt{N}|C_N - R_N| < \varepsilon] = 1 \quad .$$

Proof: Consider

$$C_N = \max_j |X_j - j/(N+1)|$$

$$= \max_j |X_j - j/N + j/N - j/(N+1)|$$

$$\leq \max_j |X_j - j/N| + 1/(N+1) = R_N + 1/(N+1) \quad .$$

Thus $C_N - R_N \leq 1/(N+1)$.

Now consider

$$R_N = \max_j |X_j - j/N|$$

$$= \max_j |X_j - j/(N+1) + j/(N+1) - j/N|$$

$$\leq \max_j |X_j - j/(N+1)| + 1/(N+1) = C_N + 1/(N+1) \quad .$$

Combining the two it is seen that

$$\sqrt{N}\,|R_N - C_N| \leq \sqrt{N}/(N+1) \quad .$$

Thus

$$P[\sqrt{N}\,|R_N - C_N| < \varepsilon] = 1$$

for all N so large that

$$\sqrt{N}\,/(N+1) < \varepsilon \quad .$$

Lemma 4: For all $\varepsilon > 0$ it will be shown that

$$\lim_{N\to\infty} P[\sqrt{N}\,|D_N - R_N| < \varepsilon] = 1 \quad .$$

Proof: Let $Q_N^- = \max_{1\leq j\leq N} |X_j - (j-1)/N|$ and note that

$$Q_N^- = \max_{1\leq j\leq N} |X_j - j/N + j/N - (j-1)/N| \leq R_N + 1/N$$

Thus

$$\sqrt{N}\,|Q_N^- - R_N| \leq 1/\sqrt{N} \quad .$$

Also

$$D_N = \max_{1 \le j \le N} \{|X_j - j/N| \;,\; |X_j - (j-1)/N|\}$$

$$= \max \{R_N \;,\; Q_N^-\} \quad .$$

Thus $|D_N - R_N| = |\max \{0, (Q_N^- - R_N)\}| \le |Q_N^- - R_N|$ and it follows that

$$\lim_{N\to\infty} P[\sqrt{N}\;|D_N - R_N| < \varepsilon] = 1$$

for all N so large that $\sqrt{N} > 1/\varepsilon$.

Lemma 5: $\lim_{N\to\infty} P[\sqrt{N}\; R_N > Z] = 2 \sum_{r=1}^{\infty} (-1)^{r-1} \exp \{2r^2 Z^2\}$.

Proof: $P[\sqrt{N}\; D_N > Z] = P[\sqrt{N}\; R_N + \sqrt{N}\;(D_N - R_N) > Z]$.

$$\lim_{N\to\infty} P[\sqrt{N}\; R_N + \sqrt{N}\;(D_N - R_N) > Z] = \lim_{N\to\infty} P[\sqrt{N}\; R_N > Z] .$$

IV. COMPUTATIONAL METHOD

The method used to evaluate the cumulative distribution function of the statistics D_N, C_N, and R_N is from the work of Durbin [Ref. 1]. A sketch of his results and the modifications necessary for the statistics under consideration are presented in this section.

"Suppose that $0 \le x_1 \le x_2 \le \ldots \le x_N \le 1$ is an ordered random sample of independent observations from a uniform (0,1) distribution. The sample distribution function is...", $F_N(x)$ as defined in section two. "Let S denote the sample path of $F_N(x)$ as x moves from 0 to 1. In this paper we consider the probability $p_n(a,b,c)$ that S lies entirely in the region R between the lines $ny = a+(n+c)x$ and $ny = -b+(n+c)x$, ($a>0$, $b>0$, $a+b>0$, and $a-c>0$)."[1]

Define:

$$H \equiv \begin{bmatrix} 1-\delta & 1 & 0 \; . \; . \; . & 0 & 0 \\ (1-\delta^2)/2! & 1 & 1 \; . \; . \; . & 0 & 0 \\ (1-\delta^3)/3! & 1/2! & 1 \; . \; . \; . & 0 & 0 \\ . & . & . \; . \; . \; . & . & . \\ . & . & . \; . \; . \; . & . & . \\ . & . & . \; . \; . \; . & . & 1 \\ \dfrac{1-\delta^P-@^P+h}{P!} & \dfrac{1-@^{P-1}}{(P-1)!} & . \; . \; . \; . & \dfrac{1-@^2}{2!} & 1-@ \end{bmatrix}$$

---

[1] Durbin, J., "The Probability That the Sample Distribution Function Lies Between Two Parallel Straight Lines." The Annals of Mathematical Statistics, V. 39, p. 398, April 1968.

$[z] \equiv$ the greatest integer in z.

$1-\delta \equiv (b-c) - [b-c]$

$1-@ \equiv (a-c) - [a-c]$

$P \equiv [b-c] + [a+c] + 1$

$h \equiv 0$ if $\delta+@ \leq 1$ and $h \equiv (\delta+@ - 1)^P$ if $\delta+@ > 1$

$q \equiv$ the $([b-c + 1],[b+1])$ element of the matrix $H^{[N+c]}$

Thus $p_n(a,b,c) = \dfrac{N!}{(N+c)^N} q$ [Ref. 1].

Lemma 6: $P[D_N \leq k/N] = p_n(k,k,0)$ [Ref.1].

Lemma 7: $P[C_N \leq k/(N+1)] = p_n(k,k+1,1)$ [Ref. 1].

Lemma 8: $P[R_N \leq k/N] = p_n(k,k+1,0)$.

Proof: Assume the sample cumulative distribution, $S_N(x)$, has stayed within the bounds shown in Figure two until it arrives at the point $(X_j, j/N)$. The next observation, X_{j+1}, must occur before X^D_{j+1} for non-rejection by the D_N statistic. For R_N to accept the sample, the next observation must occur before X^R_{j+1}. Since all steps are of height 1/N the point $(X^R_{j+1}, (j+1)/N)$ is in the region R if the lower band of the equivalent D_N statistic is shifted downward by 1/N. The resultant band is $ny = -(b+1) + nx$. If the point $(X_{j+1}, (j+1)/N)$ were to fall in the region W for either D_N or R_N the hypothesis, H_o, would be rejected. Hence the upper bound for the region R is unchanged.

The computer program used to table the distribution of C_N and R_N is presented in the rear of this thesis. As

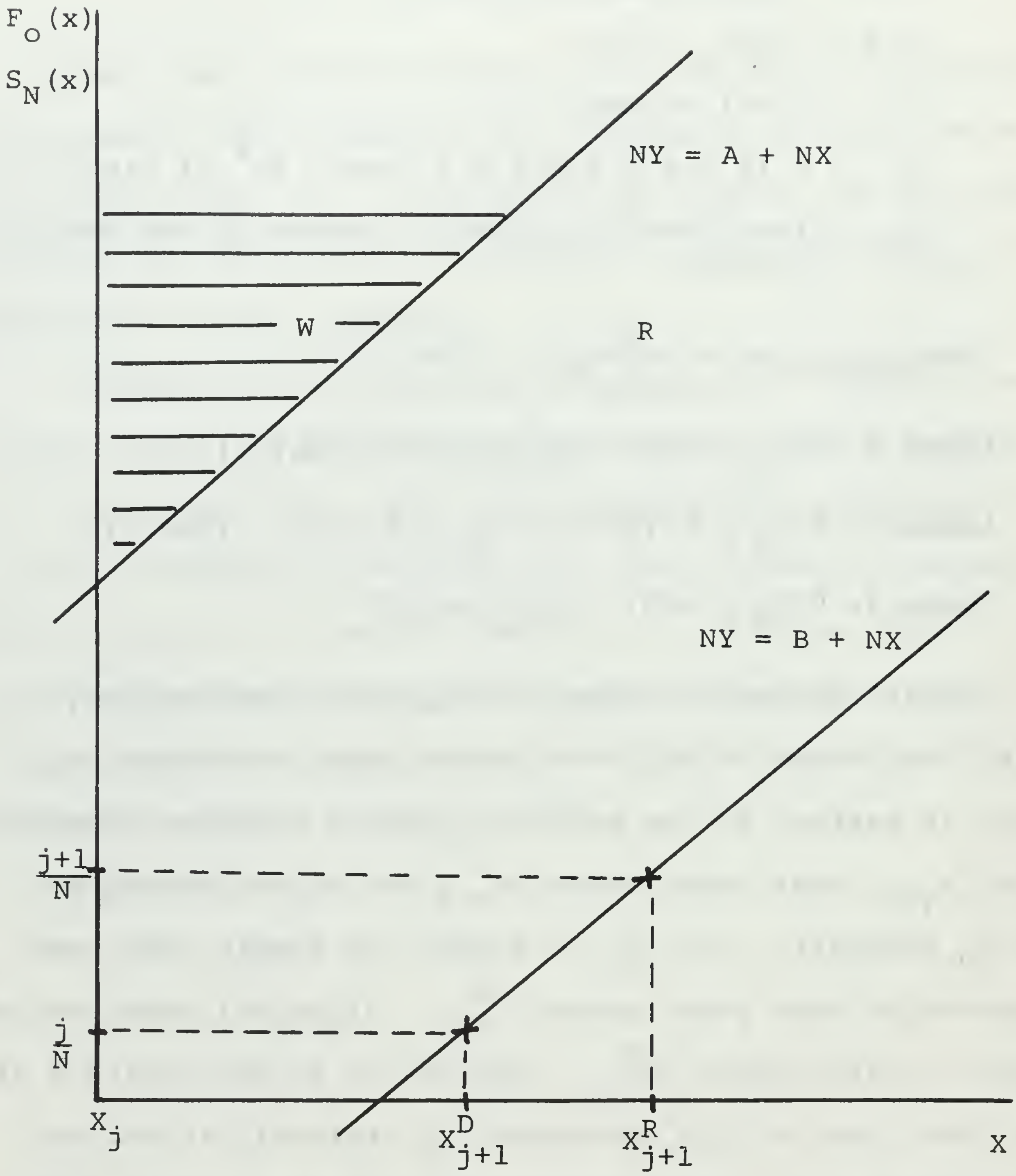

Figure 2. Lemma 8

written the parameter P is restricted to be less than or equal to forty. The subroutine DURMAT is used to keep the matrix products within the range of the IBM 360/67 used.

V. COMPARISON OF STATISTICS, REMARKS, AND CONCLUSIONS

The distribution of the statistics C_N and R_N are presented for the first time in Appendix A and B. Some critical points for various levels of significance are presented in Appendix C for D_N, C_N, and R_N. It is seen for sample sizes greater than sixty all three statistics have essentially the same critical points for alpha = .20, .10, .05, .025, and .01.

From lemma two it is known that R_N is stochastically smaller than D_N and Appendix C indicates that C_N is also stochastically smaller than D_N for the alpha values shown under H_o. Such results are important in developing smaller confidence bands for distributions [Ref. 7].

It has been noted that the asymptotic formula of Kolmogorov is correct (to three decimal places) if $N>60$ for all three statistics. The tables indicate that the limiting distribution is reached by D_N earlier (i.e., $N\geq 30$) for the alpha values listed. It can be seen in Appendix C that for alpha $\geq .05$ the distribution of D_N approaches the asymtote from above whereas both C_N and R_N approach it from below for all listed values of alpha. Also, C_N and R_N are in essential agreement for sample size greater than fifteen.

Other statistics can be investigated empirically and analytically using Durbin's paper. An interesting area for investigation would be to relate the power functions of the test statistics to the parameters (a,b,c).

APPENDIX A.

$$P\left(C_N < K/(N+1) \right)$$

K \ N =	5	10	15	20	25	30	35
1	.27006	.03550	.00417	.00047	.00005	.0	.0
2	.83719	.46822	.23292	.11016	.05062	.02283	.01016
3	.98666	.83796	.63607	.45822	.32060	.22015	.14921
4	.99999	.96862	.87604	.75491	.63168	.51860	.42018
5	1.0	.99626	.96838	.91124	.83568	.75243	.66858
6		.99974	.99403	.97403	.93814	.89000	.83417
7		.99998	.99917	.99390	.98034	.95747	.92640
8		.99999	.99991	.99885	.99474	.98570	.97099
9		.99999	.99998	.99981	.99881	.99582	.98984
10		1.0	.99998	.99995	.99975	.99892	.99683
11						.99973	.99910
12						.99991	.99974

K \ N =	40	45	50	55	60	65	70
3	.10016	.06674	.04420	.02914	.01913	.01252	.00817
4	.33714	.26850	.21257	.16749	.13154	.10281	.08019
5	.58829	.51376	.44601	.38532	.33157	.28436	.24317
6	.77448	.71374	.65384	.59608	.54125	.48980	.44197
7	.88906	.84738	.80305	.75745	.71160	.66627	.62203
8	.95078	.92581	.89700	.86533	.83164	.79665	.76098
9	.98030	.96708	.95073	.93062	.90829	.88387	.85782
10	.99288	.98666	.97796	.96680	.95332	.93771	.92024
11	.99766	.98666	.99097	.98524	.97779	.96860	.95777
12	.99929	.99505	.99657	.99389	.99011	.98512	.97888
13				.99762	.99587	.99335	.99001
14				.99911	.99835	.99718	.99551

K \ N =	75	80	85	90	95	100
3	.00532	.00345	.00224	.00144		
4	.06238	.04843	.03753	.02903		
5	.20743	.17657	.15001	.12723		
6	.39782	.35731	.32030	.28665		
7	.57929	.53832	.49928	.46229	.42740	.39460
8	.72512	.68946	.65430	.61988	.58639	.55395
9	.83060	.80257	.77402	.74524	.71647	.68787
10	.90121	.88087	.85947	.83725	.81443	.79119
11	.94542	.93169	.91671	.90068	.88374	.86605
12	.97141	.96275	.95292	.94204	.93020	.91749
13	.98579	.98067	.97460	.96764	.95982	
14	.99328	.99044	.98691	.98271	.97781	

APPENDIX B.

$P(R_N < K/N)$

K	N = 5	10	15	20	25	30
1	.27854	.03814	.00455	.00051	.00006	
2	.83040	.47338	.23874	.11374	.05253	.02377
3	.98464	.83505	.63716	.46151	.32425	.22334
4	.99970	.96661	.87390	.75422	.63247	.52053
5	1.0	.99582	.96699	.90962	.83454	.75201
6		.99970	.99358	.97307	.93696	.88891
7		.99998	.99909	.99354	.97969	.95663
8		.99999	.99989	.99876	.99448	.98526
9		.99999	.99997	.99980	.99873	.99564
10		1.0	.99998	.99996	.99974	.99887

K	N = 35	40	45	50	55	60
6	.83336	.77405	.71369	.65414	.59666	.54205
7	.92551	.88824	.84672	.80260	.75721	.71152
8	.97039	.95010	.92511	.89635	.86476	.83117
9	.98954	.97988	.96658	.94983	.93006	.90773
10	.99671	.99268	.98638	.97761	.96639	.95287
11	.99906	.99758	.99492	.99078	.98499	.97748
12	.99974	.99925	.99825	.99648	.99376	.98993
13	.99990	.99976	.99942	.99873	.99756	.99577
14	.99994	.99990	.99979	.99955	.99908	.99831

K	N = 65	70	75	80	85	90
6	.49076	.44302	.39892	.35854	.32142	.28773
7	.66645	.62239	.57979	.53893	.49999	.46306
8	.79632	.76078	.72506	.68953	.65448	.62016
9	.88337	.85739	.83024	.80229	.77383	.74513
10	.93725	.91979	.90077	.88047	.85911	.83694
11	.96828	.95714	.94504	.93130	.91633	.90031
12	.98492	.97864	.97114	.96244	.95261	.94171
13	.99324	.98986	.98562	.98046	.97438	.96739
14	.99713	.99543	.99318	.99031	.98677	.98253

APPENDIX C. ASYMPTOTIC COMPARISON OF D_N, C_N, AND R_N

ALPHA	N=	10	15	20	25	30	35
.20	A	.322	.266	.231	.210	.192	.181
	D	.338	.284	.239	.212	.196	.181
	C	.292	.246	.205	.193	.178	.166
	R	.309	.245	.205	.193	.178	.166
.10	A	.368	.304	.264	.238	.221	.206
	D	.386	.316	.273	.244	.223	.206
	C	.348	.284	.246	.226	.205	.195
	R	.342	.284	.246	.226	.205	.195
.05	A	.409	.338	.294	.264	.242	.230
	D	.431	.352	.304	.272	.248	.230
	C	.386	.321	.281	.251	.230	.217
	R	.386	.321	.281	.251	.230	.217
.025	A	.463	.378	.327	.293	.268	.248
	D	.453	.364	.324	.292	.268	.248
	C	.425	.351	.300	.275	.254	.244
	R	.428	.351	.300	.275	.254	.244
.01	A	.516	.423	.365	.326	.298	.276
	D	.487	.404	.352	.321	.293	.276
	C	.478	.390	.343	.308	.283	.258
	R	.481	.390	.343	.308	.283	.258

ALPHA	N=	40	45	50	60	N>60
.20	A	.169	.160	.157	.138	
	D	.169	.160	.157	.138	
	C	.156	.148	.140	.134	$1.07/\sqrt{N}$
	R	.156	.148	.140	.134	
.10	A	.193	.182	.173	.158	
	D	.193	.182	.173	.158	
	C	.179	.170	.161	.155	$1.22/\sqrt{N}$
	R	.179	.170	.161	.155	
.05	A	.216	.203	.194	.176	
	D	.216	.203	.194	.176	
	C	.200	.190	.180	.173	$1.36/\sqrt{N}$
	R	.200	.190	.180	.173	
.025	A	.231	.218	.207		
	D	.231	.218	.207		
	C	.222	.209	.200		$1.43/\sqrt{N}$
	R	.222	.209	.200		
.01	A	.258	.243			
	D	.258	.243			
	C	.249	.243			$1.63/\sqrt{N}$
	R	.249	.243			

A = ASYMPTOTIC ESTIMATION
D = D_N
C = C_N
R = R_N

COMPUTER OUTPUT

```
                 SAMPLE PROGRAM
          THE D STATISTIC FOR N=10
 IA=   1 IB=   1 IC=   0 NNN= 10 PROB. = 0.0003628
 IA=   2 IB=   2 IC=   0 NNN= 10 PROB. = 0.2512797
 IA=   3 IB=   3 IC=   0 NNN= 10 PROB. = 0.7294613
 IA=   4 IB=   4 IC=   0 NNN= 10 PROB. = 0.9410061
 IA=   5 IB=   5 IC=   0 NNN= 10 PROB. = 0.9922168
 IA=   6 IB=   6 IC=   0 NNN= 10 PROB. = 0.9994260
 IA=   7 IB=   7 IC=   0 NNN= 10 PROB. = 0.9999743
 IA=   8 IB=   8 IC=   0 NNN= 10 PROB. = 0.9999936
 IA=   9 IB=   9 IC=   0 NNN= 10 PROB. = 0.9999938

   OUT OF RANGE IA= 10 IB= 10 IC=   0 NNN= 10

 THE ABOVE IS AN EXAMPLE OF ONE OF THE ERROR MESSAGES
```

COMPUTER PROGRAM

```
    1 FORMAT(16X,'SAMPLE PROGRAM',/)
    2 FORMAT(10X,'THE D STATISTIC FOR N=10',/)
    3 FORMAT(1X,'THE ABOVE IS AN EXAMPLE OF',
     1' ONE OF THE ERROR MESSAGES.'/)
C
C
      WRITE(6,1)
      WRITE(6,2)
      IC=0
      NNN=10
      DO 10 IA=1,10
   10 CALL DURBIN(IA,IA,IC,NNN)
      WRITE(6,3)
      STOP
      END

      SUBROUTINE DURBIN(IA,IB,IC,NNN)
C
C          THIS SUBROUTINE COMPUTES THE PROBABILITY THAT A
C     RANDOM SAMPLE OF SIZE NNN DRAWN FROM A UNIFORM(0,1)
C     RANDOM VARIABLE AND PLOTTED AS SHOWN IN SECTION
C     FOUR OF THIS THESIS WILL STAY WITHIN THE REGION, R,
C     BETWEEN THE LINES,
C       NNN*Y=NNN*IA+(NNN+IC)*X
C        NNN*Y= -NNN*IB+(NNN+IC)*X
C
C
C     THE RESTRICTIONS ARE
C     IA, IB, AND IC MUST BE INTEGER VARIABLES
C     IC.GE.0
C     IA.GT.0
C     IB.GT.0
C     IA+IC.GT.0
C     IB-IC.GT.0
C
    3 FORMAT(///,2X,' OUT OF RANGE IA=',I3,' IB=',' IC= ',
     1I3,' NNN= ',I3,//)
    4 FORMAT(//,' IP.GT.NNN IA=',I3,' IB=',I3,
     1' IC=',I3,' IP= ',I3,' NNN= ',I3)
    5 FORMAT(' IA=',I3,' IB=',I3,' IC=',I3,' NNN=',I3,
     1 ' PROB. =   ',F9.8)
C
C
      DIMENSION H(1600),HSTORE(1600),SAVE1(1600),SAVE2(1600)
      DATA H/1600*0.0/
      DATA HSTORE/1600*0.0/
      DATA SAVE1/1600*0.0/
      DATA SAVE2/1600*0.0/
      DATA JPH/0/
      IF(IA.LE.0) GO TO 10000
      IF(IB.LE.0) GO TO 10000
      IF((IA+IC).LE.0) GO TO 10000
      IF((IB-IC).LE.0) GO TO 10000
      IF(JPH.GT.0) GO TO 100
      CALL DURMAT(HSTORE,40)
      JPH=100
  100 IP=IA+IB+1
      IF(IP.GT.40) GO TO 10000
      IJ=IB*IP+IB-IC+1
      IF(((IB+IC).GE.NNN).AND.((IA+IC).GE.NNN)) GO TO 11000
      NSIZE=IP*IP
      JGOT=0
      DO 500 I=1,IP
      JGET=(I-1)*40
      DO 500 J=1,IP
      JGET=JGET+1
      JGOT=JGOT+1
```

```
  500 H(JGOT)=HSTORE(JGET)
      DO 600 J=1,IP
      M=J*IP
  600 H(M)=0.0
      DO 700 J=1,IP
      M=(J-1)*IP+J
  700 SAVE2(M)=1.0
      ICK=NNN+IC
      FACT=REDUCE(NNN,IC,1)
  750 CONTINUE
      DO 800 J=1,NSIZE
  800 SAVE2(J)=SAVE2(J)*FACT
      CALL GMPRD(SAVE2,H,SAVE1,IP,IP,IP)
      IPATH=1
      ICK=ICK-1
      IF(ICK.LE.0) GO TO 2000
      FACT=REDUCE(NNN,IC,2)
      DO 900 J=1,NSIZE
  900 SAVE1(J)=SAVE1(J)*FACT
      CALL GMPRD(SAVE1,H,SAVE2,IP,IP,IP)
      IPATH=2
      ICK=ICK-1
      IF(ICK.LE.0) GO TO 2000
      FACT=REDUCE(NNN,IC,2)
      GO TO 750
 2000 GO TO (2010,2100), IPATH
 2010 WRITE(6,5)IA,IB,IC,NNN,SAVE1(IJ)
      GO TO 2500
 2100 WRITE(6,5)IA,IB,IC,NNN,SAVE2(IJ)
 2500 DO 2600 I=1,NSIZE
 2600 SAVE2(I)=0.0
      RETURN
10000 CONTINUE
      WRITE(6,3)IA,IB,IC,NNN
      RETURN
11000 WRITE(6,4)IA,IB,IC,IP,NNN
      RETURN
      END

      FUNCTION REDUCE(NNN,IC,IENTER)
C
C          THIS SUBROUTINE COMPUTES THE TRANSITION MATRIX
C     H OF SECTION FOUR OF THIS THESIS FOR INTEGER VALUES
C     OF IA, IB, AND IC.
C
C
      GO TO (1,100),IENTER
    1 CONTINUE
      Z=NNN+IC
      W=NNN
      ICK=NNN-1
      Y=0.0
      IPATH=2
      SAVE=W/Z
      W=W-1.
      REDUCE=SAVE
      RETURN
  100 IF(ICK.LE.0) GO TO 1000
      GO TO(200,300),IPATH
  200 IPATH=2
      ICK=ICK-1
      SAVE=W/Z
      W=W-1.0
      REDUCE=SAVE
      RETURN
  300 IPATH=1
      Y=Y+1.0
      ICK=ICK-1
      REDUCE=Y/Z
      RETURN
 1000 REDUCE=1.0
```

```
      RETURN
      END

      SUBROUTINE GMPRD(A,B,R,N,M,L)
C             THIS SUBROUTINE COMPUTES THE MATRIX PRODUCT OF
C     A*B AND STORES THE RESULTS IN R.
      DIMENSION A(1),B(1),R(1)
      IR=0
      IK=-M
      DO 10 K=1,L
      IK=IK+M
      DO 10 J=1,N
      IR=IR+1
      JI=J-N
      IB=IK
      R(IR)=0.0
      DO 10 I=1,M
      JI=JI+N
      IB=IB+1
   10 R(IR)=R(IR)+A(JI)*B(IB)
      RETURN
      END

      SUBROUTINE DURMAT(DUR,JDUR)
C
C     THIS SUBROUTINE PROVIDES THE SCALE FACTOR OF
C     P(IA,IB,IC) OF SECTION FOUR IN THE FOLLOWING ORDER,
C     N/N,1/N,(N-1)/N,2/N,...
C
      DIMENSION DUR(1)
      NNN=JDUR*JDUR
      DO 10 I=1,NNN
   10 DUR(I)=0.0
      IJ=JDUR
      DO 200 I=2,JDUR
      IJ=IJ+I-1
      DUR(IJ)=1.0
      IJ=IJ+1
      DUR(IJ)=1.0
      IF(IJ.GE.NNN) GO TO 300
      NN=I+1
      Z=2.0
      DO 100 IK=NN,JDUR
      IJ=IJ+1
      DUR(IJ)=DUR(IJ-1)*(1.0/Z)
  100 Z=Z+1.0
  200 CONTINUE
  300 RETURN
      END
```

BIBLIOGRAPHY

1. Durbin, J. "The Probability That the Sample Distribution Function Lies Between Two Parallel Straight Lines," The Annals of Mathematical Statistics, 39:398-411, April 1968.

2. Kolmogorov, A. "Sulla Determinazione Empirica di una Legge di Distribution," Giornale dell'Instituto degi Attuari, 4:1-11, 1933.

3. Massey, F. J. "A Note on the Estimation of a Distribution Function by Confidence Limits," The Annals of Mathematical Statistics, 21:116-119, March, 1950.

4. Massey, F. J. "The Kolmogorov-Smirnov Test for Goodness of Fit," Journal of the American Statistical Association, 46:68-78, March, 1951.

5. Pyke, R. "The Supremum and Infimum of the Poisson Process," The Annals of Mathematical Statistics, 30:568-576, June, 1959.

6. Smirnov, H. "Sur les Ecarts de la Courbe de Distribution Empirique," Recueil Mathematique, 6:3-26, 1939.

7. Wald, A. and Wolfowitz, J. "Confidence Limits for Continuous Distribution Function," The Annals of Mathematical Statistics, 10:105-118, June, 1939.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandriz, Virginia 22314	20
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Chief of Naval Operations (OP-96) Department of the Navy Washington, D. C. 20350	1
4.	Professor J. R. Borsting, Code 55 Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	2
5.	Associate Professor R. R. Read, Code 55Re Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	2
6.	Assistant Professor D. R. Barr, Code 55Bn Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	1
7.	Professor F. J. Massey University of California at Los Angeles School of Public Health Los Angeles, California 90024	1
8.	Professor Ronald Pyke University of Washington Seattle, Washington 98105	1
9.	Professor J. Durbin London School of Economics Huston Street London, England WC2	1
10.	LT Jasper Paul Hendren 1538 Sunset Drive New Albany, Indiana 47150	1

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D		
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)		
1. ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP	
3. REPORT TITLE A COMPARISON OF SOME TEST STATISTICS OF THE KOLMOGOROV TYPE		
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Master's Thesis; October 1969		
5. AUTHOR(S) (First name, middle initial, last name) Jasper Paul Hendren, Lieutenant, United States Navy		
6. REPORT DATE October 1969	7a. TOTAL NO. OF PAGES 30	7b. NO. OF REFS 7
8a. CONTRACT OR GRANT NO. b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
10. DISTRIBUTION STATEMENT Distribution of this document is unlimited.		
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Naval Postgraduate School Monterey, California 93940	

13. ABSTRACT

It is natural to associate optimal estimation and optimal statistical testing. In this paper continuous function estimation of the cumulative distribution is used to define two test statistics that compete with the Kolmogorov D_N statistic. The first statistic, C_N, is attributed to Pyke and the second, R_N, is obtained by polygonalizing the sample distribution function. It is known that both are asymptotically equivalent to the Kolmogorov statistic. Using the methods of J. Durbin, the small sample distributions are tabled as well as the critical points for significance levels of .20, .10, .05, .025, and .01. It is shown that R_N is stochastically smaller than D_N, and it appears that C_N is also smaller than D_N.

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0101-807-6811

UNCLASSIFIED
Security Classification

A-31408

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
STATISTICAL TEST						
KOLMOGOROV STATISTIC						
PYKE STATISTIC						
NON-PARAMETRIC STATISTICS